Original Article

A Homogeneous Ensemble of Robust Pre-defined Neural Network Enables Automated Annotation of Human Embryo Morphokinetics

Gunawan B. Danardono [1, 2*], Alva Erwin [1, 2], James Purnama [2], Nining Handayani [1, 3], Arie A. Polim [1, 3, 4], Arief Boediono [1, 3, 5], Ivan Sini [1, 3]

1- IRSI Research and Training Center, Jakarta, Indonesia
2- Faculty of Engineering and Information Technology, Swiss German University, Tangerang, Indonesia
3- Morula IVF Jakarta Clinic, Jakarta, Indonesia
4- Department of Obstetrics and Gynecology, School of Medicine and Health Sciences, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
5- Department of Anatomy, Physiology and Pharmacology, IPB University, Bogor, Indonesia

* *Corresponding Author:*
Gunawan B. Danardono,
IRSI Research and Training
Center, Jakarta, Indonesia
E-mail:
gunawanbondan@gmail.com

Abstract

Background: The purpose of the current study was to reduce the risk of human bias in assessing embryos by automatically annotating embryonic development based on their morphological changes at specified time-points with convolutional neural network (CNN) and artificial intelligence (AI).

Methods: Time-lapse videos of embryo development were manually annotated by the embryologist and extracted for use as a supervised dataset, where the data were split into 14 unique classifications based on morphological differences. A compilation of homogeneous pre-trained CNN models obtained via TensorFlow Hub was tested with various hyperparameters on a controlled environment using transfer learning to create a new model. Subsequently, the performances of the AI models in correctly annotating embryo morphologies within the 14 designated classifications were compared with a collection of AI models with different built-in configurations so as to derive a model with the highest accuracy.

Results: Eventually, an AI model with a specific configuration and an accuracy score of 67.68% was obtained, capable of predicting the embryo developmental stages (t1, t2, t3, t4, t5, t6, t7, t8, t9+, tCompaction, tM, tSB, tB, tEB).

Conclusion: Currently, the technology and research of artificial intelligence and machine learning in the medical field have significantly and continuingly progressed in an effort to develop computer-assisted technology which could potentially increase the efficiency and accuracy of medical personnel's performance. Nonetheless, building AI models with larger data is required to properly increase AI model reliability.

Keywords: Artificial intelligence, Automation, Computer-assisted image processing, Embryonic development, In vitro fertilization, Machine learning, Neural networks.

To cite this article: Danardono GB, Erwin A, Purnama J, Handayani N, Polim AA, Boediono A, et al. A Homogeneous Ensemble of Robust Pre-defined Neural Network Enables Automated Annotation of Human Embryo Morphokinetics. J Reprod Infertil. 2022;23(4):250-256. https://doi.org/10.18502/jri.v23i4.10809.

Introduction

To reduce and minimize human errors during IVF, developing an automated system to assist embryologists in the laboratory has become eminent. Conventional embryo grading or annotation of embryo morphokinetics is conducted through manual observation of the embryo culture by an embryologist which is energy- and time-consuming (1). Machine learning (ML) could potentially serve as a solution, especially with the recent advancement in neural networks

and publicly available learning resources. In vitro fertilization as one of the effective treatments for infertility has become a trending subject for artificial intelligence (AI) and ML experiments. Efficiency of cell annotation and separating cells based on morphokinetic differences reduce the challenge of subjective assessment of embryo cells. Research in AI and ML in medical field had been established by a number of researchers who attempted the application of neural network system for binary image classification such as euploid prediction (2-4) and livebirth prediction (5-7), and for multiclass classification such as classification of embryo development stage (8-11) and embryo grading (12-14).

Typically, an embryo culture is maintained for 5 days from the fertilization stage up to the blastocyst stage. Through the time-lapse technology, the entire embryo development process could be recorded and compressed into a time-lapse video which becomes the main reference for observation. The embryologists normally require 5 *min* to observe and annotate a single embryo development cycle (15). Automation of the classification process would therefore accelerate the embryo annotation procedure and reduce the subjectivity of individual embryologist assessment. In this study, utilization of convolutional neural network (CNN) as part of deep learning algorithm was initiated to construct an AI model capable of classifying embryo developmental stages based on annotated datasets from expert embryologists. According to VerMilyea et al., an artificial model is 24.7% more accurate than the embryologists in meticulously assessing embryo morphology on day 5 and predicting clinical pregnancy (13).

The embryo developmental stages are distinguished by specific events in the process of fertilization, cleavage, morula formation up until the blastulation. Embryo morphokinetics refers to the time-associated transformation of the embryo as the cells go through cell division or cell splitting.

Upon intracytoplasmic sperm injection (ICSI), a procedure of injecting a selected sperm cell directly into an ooplasm (16), the embryo culture/morphokinetics begins (denoted as time zero) and is maintained for 5 to 6 days until the critical blastocyst stage is reached.

Methods

This study utilized images of embryos that were cultured in MIRI® time-lapse incubators (37°C, 5% CO_2, and 5% O_2), a closed incubator system that permits the record-keeping of uninterrupted time-lapse videos (Single data center: Morula IVF Jakarta Clinic, Jakarta, Indonesia). Images were captured in a sequence of time, then merged and compiled to create a time-lapse video of the embryo morphokinetics. The time-lapse incubator can accommodate multiple embryos at a time while maintaining a sterile environment for the embryo growth. Time-lapse technology has elicited an increase in the probability of clinical pregnancy and embryo implantation in IVF (17). The annotated time-lapse video was used as a dataset in this prospective study with 163 embryo cycles which consisted of the selected time-points: t1, t2, t3, t4, t5, t6, t7, t8, t9+, tcompaction, tM, tSB, tB, and tEB. While maintaining the integrity of the dataset, imbalanced dataset could not be avoided. Instantaneous cell split between two time-points periodically occur resulting in rapid morphological changes; hence, the dataset of embryos in t3, t5, t6 and t7 are smaller in quantity (18). Table 1 summarizes the dataset used in this study which amounts to 15.831 embryo images at different developmental stages.

Convolutional neural network is a feed-forward neural network (FFNN) which uses a mathematical linear equation between matrices called the convolution (19). A basic CNN would require convolutional, pooling, and fully-connected layers to represent its characteristics (20). A supervised dataset is fed into the CNN with input images that

Table 1. Embryo dataset

Time point	T1	T2	T3	T4	T5	T6	T7	T8	T9+	Compaction	Morula	Early blastocyst	Blastocyst	Expanded blastocyst
Image no.	3189	1388	348	1540	323	378	445	-	-	-	-	-	-	-
Image no.	-	-	-	-	-	-	-	1575	1197	1135	1668	754	930	961

have been grouped based on the header or pre-classified classes as part of the original images. The last layer would consist of a 1×1 dimensional node with multiple depths; depth in this context depicts the possible number of classifications and simplification was done through a nonlinear function (21).

To produce a CNN model, TensorFlow and Keras library were implemented since both are open-source libraries created by Google and community developers. TensorFlow employs graphics processing unit (GPU) as one of the computational devices to accelerate data clustering and training process (22). To represent a TensorFlow for model algorithms, dataflow graphs utilize layers as mathematical matrix operators and each type of layer produces a different mathematical output (23). TensorFlow is designed to provide an efficient memory and computation, a stable numerical system, and maintain an idiom to differentiate its ecosystem (24). Alternatively, Keras is particularly built for easy and efficient experimentation besides faster results in research (25). Keras itself is a high-level application programming interface (API) which can serve neural network purposes.

TensorFlow and Keras have published multiple open-source pre-trained models on TensorFlow Hub that could be used for transfer learning using a newly introduced dataset. The transfer learning allows leveraging feature representations from pre-trained models with pre-defined layers and altering the end point of the model into a desired classification. Each individual pre-trained model is unique; for instance, inception_v3 has 313 layers, ResNet has 566 layers while efficientnet_b6 has 669 layers (22). In transfer learning, a pre-trained model with all its layers is considered as a singular layer (one entity) which allows users to use either sequential or functional API. Figure 1 shows the architectural differences between functional and sequential API. Functional API provides flexibility in creating multiple parallel layers, while sequential API exhibits linear layers. To achieve a robust AI model, this study combined pre-trained model selection, data augmentation, and hyperparameter selection. Pre-trained model selection was conducted through trials of multiple different model architectures trained for 100 learning steps or epochs. Data augmentation consisted of random image rotation and image flip in the dataset, and hyperparameter selection comprised testing multiple optimizers and learning rates. Configuration selection was performed in a shorter training time (30 learning steps or epochs) compared to the final model built (200 learning steps or epochs).
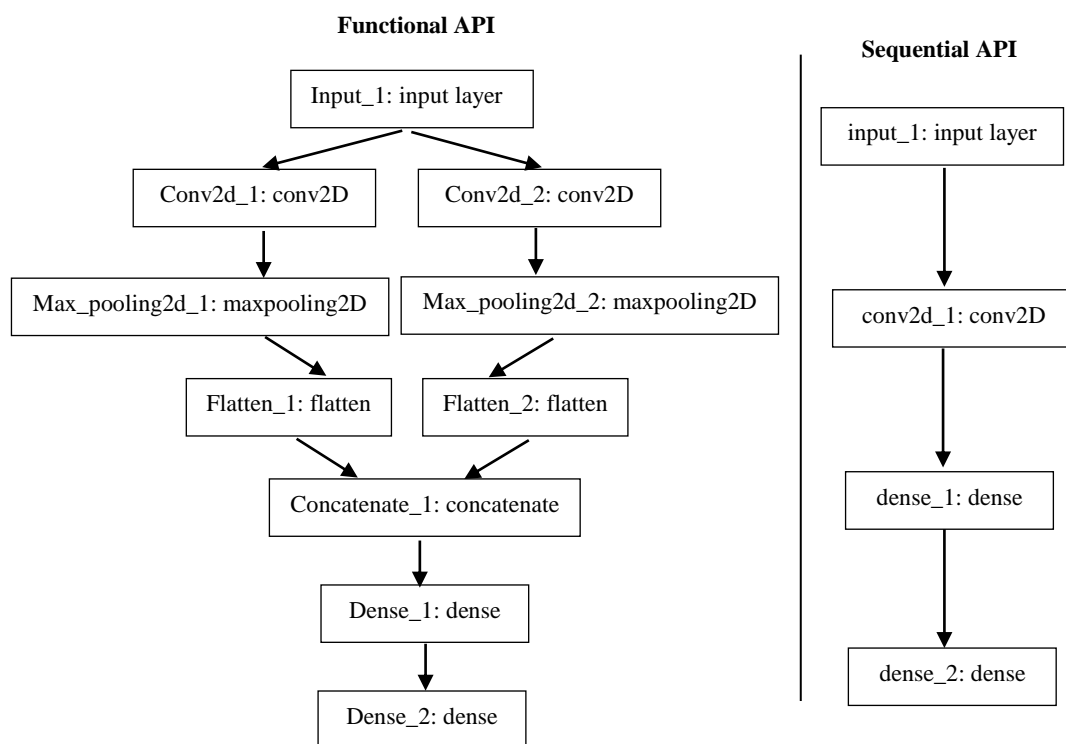
Figure 1. The comparison of functional and sequential API

Ethical committee approval: This study was approved by the local research ethics committee in the Faculty of Medicine, Universitas Indonesia, Jakarta (number: KET-351/UN2.F1/ETIK/PPM.00.02/2020).

Results

In this study, multiple model architectures with different hyperparameters were compared with/ without data augmentation to identify the most suitable architecture which best achieves the functions of embryonic cell detection and annotation. Retrospective IVF patient data on successful embryo development up to the expanded blastocyst were obtained for this study. The dataset consisted of image records of the embryonic cells after sperm injection, at the T1 to T9 stage, compaction, morula, and blastocyst stage. Figure 2 shows the sequence of the embryonic development and different morphokinetic parameters captured using the time-lapse incubator camera. These images had undergone the pre-processing steps of cropping, thresholding, erosion, and dilation prior to extraction from the original time-lapse videos. Six pre-trained models were used in this study with multiple parameters to identify which model has the highest accuracy. Sequential API was used for the comparison because of its capability to gain access to TensorFlow Hub and linear model

Figure 2. Embryonic development events

layer computation. The pre-trained model yielded an x amount of output which was then narrowed to a specific classification using a dense layer. The dense layer serves to limit the outputs to the desired classification. The pre-trained model selection was conducted using the following parameters: 100 epochs, Adam optimizer, and 5E-03 learning rate without data augmentation. Table 2 shows the output accuracy of the different pre-trained models with similar parameters. Pre-trained models with the highest accuracy would subsequently be used for hyperparameter selection and final model training.

Evidently, some architectures performed better than the others. Each pre-trained model has a different layer approach which influences its performance. The purpose of the layers is to update the model's weight for each node; more layers presumably yield a better performance. In this study, efficientnet_b6 architecture performance was proven to be superior and it utilized more layers than the other models and thus was used for the hyperparameter selection. The trial for hyperparameter selection was conducted over 30 epochs to expedite the initial probability results that would become a benchmark for downstream training steps.

Optimizer, learning rate, and data augmentation were factors that were tested using commonly used hyperparameters and similar configuration attempts. The optimizers were derived from the Keras library. The learning rate used a sequence order of number selection. The data augmentation was conducted to potentially increase the model performance by using two randomly selected settings. Learning rate determines the model capability to adapt with training progress. A small value of learning rate indicates a slow training progress while a high learning rate value would result in different loss functions. The calculation for each learning rate differs for each optimizer, since each uses a different set of mathematical approaches to
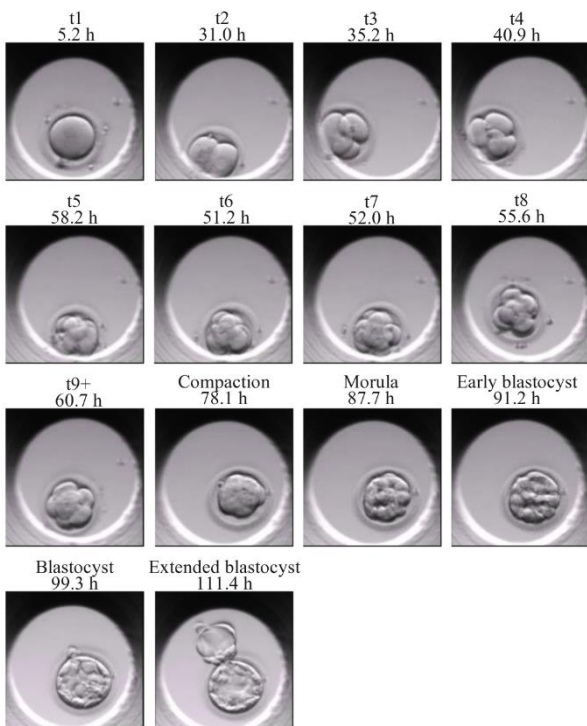
Table 2. Performances of pre-trained models

No	Model	Accuracy (%)
1	Inception_v3	40-50
2	Efficientnet_b3	50-60
3	Mobilenet_v1_100_224	35-45
4	Mobilenet_v3_large_075_224	55-60
5	Efficientnet_b6	55-60
6	Resnet_v2_152	55-60

comply with their individual goals. The combination of learning rate and optimizer yielded better test results as shown in table 3.

Implementing random rotation and random flip on the embryo image dataset did not provide any improvement in the model accuracy. In fact, it generated a slightly lower accuracy compared to the original dateset, 58.40% and 59.86%, respectively. The data augmentation process produced 32 different versions of an individual image with the same dimension. Random zoom data augmentation or partial image entry was excluded and considered unnecessary because the original input dataset had no incomplete frames and all embryo images were confined in one frame. The images were gathered from a single data center which generates equal light intensity of images, hence random augmentation of light intensity was excluded. Additionally, all images collected for the study were clear images without any distortions, thus random distortion of image or noise augmen-

tation was excluded. Nonetheless, model training using the original image dataset is assumed to produce better outcomes than using the altered images.

Consequently, the model was trained using the efficientnet_b6 architecture for 200 epochs to finalize the training sequence. Training for 200 epochs was determined to expectedly improve the model performance. In general, 200 epochs were deemed sufficient for the training and that a higher epoch value would not contribute to a significant difference.

Table 4 shows the results of a 200 epoch model training without data augmentation. Models with the Adam AMSGrad optimizer and the learning rate of 1E-03 yielded the highest performance compared to other hyperparameter selection models.

Discussion

In this study, Keras and TensorFlow were used

Table 3. Learning rate comparison

No	Optimizer	Learning rate	The highest training accuracy (%)
1	Adamax	5E-03	58.78
2	Adadelta	5E-02	41.69
3	RMSprop	1E-03	57.91
4	Nadam	5E-03	58.49
5	Adagrad	5E-02	57.61
6	Adam	5E-03	60.83
7	FTRL	5E-02	58.49
8	Adam-AMSGrad	1E-03	61.42
9	SGD	1E-02	58.88

Table 4. Model finalization

No	Optimizer	Learning rate	Test		Validation	
			Accuracy (%)	Loss	Accuracy (%)	Loss
1	Adamax	5E-03	67.58	1.40	55.76	1.67
2	Adadelta	5E-02	57.52	1.63	53.42	1.72
3	RMSprop	1E-03	67.29	1.41	57.03	1.68
4	Nadam	5E-03	66.99	1.46	56.05	1.75
5	Adagrad	5E-02	65.92	1.43	56.35	1.66
6	Adam	5E-03	65.63	1.50	55.96	1.75
7	FTRL	5E-02	65.92	1.41	57.03	1.62
8	Adam AMSGrad	1E-03	67.68	1.40	57.71	1.65
9	SGD	1E-02	66.24	1.43	55.47	1.66

to perform transfer learning and the performances of several pre-trained models with various configurations were compared. Furthermore, hyperparameter selection was attempted to identify models with the most optimized performance. Pre-trained models with the efficientnet_b6 architecture yielded the highest accuracy compared to other pre-trained model architectures. Ultimately, the hyperparameters were compared with different configurations and the model with the optimizer Adam AMSGrad and 1E-03 learning rate was determined to produce the highest accuracy of 67.68% without the use of data augmentation. Each pre-trained model architecture exhibited unique layers, even for similar models with different iterations. The complex 660 layers of the efficientnet_b6 architecture in this study overpowered the performances of other models. The differences in the model architectures and input data size therefore had an effect on the model performances. Previous studies on embryo development classification (8-11) utilized the early stages of embryo development up to t4, t4+, t5, t4+, respectively. The novelty of the current study was the added time-points in the morphokinetic parameters, summing up to 14 kinetic stage classifications (up to the expanded blastocyst stage), discernable by the AI model.

Conclusion

The advanced technology such as CNN is capable for image classification to support and improve the decision-making process of medical personnel. Such technology would provide embryologists with a benchmark to annotate embryos at t1 up to tEB, instead of relying solely on manual observation. Moreover, the AI model constructed in this study could significantly be improved through training with a larger dataset as data is a crucial factor that determines a model's performance in the field of machine learning and artificial intelligence.

Acknowledgement

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Kanakasabapathy MK, Thirumalaraju P, Bormann CL, Kandula H, Dimitriadis I, Souter I, et al. Development and evaluation of inexpensive automated deep learning-based imaging systems for embryology. Lab Chip. 2019;19(24):4139-45.

2. Chavez-Badiola A, Flores-Saiffe-Farías A, Mendizabal-Ruiz G, Drakeley AJ, Cohen J. Embryo ranking intelligent classification algorithm (ERICA): artificial intelligence clinical assistant predicting embryo ploidy and implantation. Reprode Biomed Online. 2020;41(4):585-93.

3. Lee CI, Su YR, Chen CH, Chang TA, Kuo EE, Zheng WL, et al. End-to-end deep learning for recognition of ploidy status using time-lapse videos. J Assist Reprod Genet. 2021;38(7):1655-63.

4. Huang B, Tan W, Li Z, Jin L. An artificial intelligence model (euploid prediction algorithm) can predict embryo ploidy status based on time-lapse data. Reprod Biol Endocrinol. 2021;19(1):185.

5. Kanakasabapathy MK, Thirumalaraju P, Bormann CL, Gupta R, Pooniwala R, Kandula H, et al. Deep learning mediated single time-point image-based prediction of embryo developmental outcome at the cleavage stage. arXiv preprint arXiv:2006.08346. 2020 May 21.

6. Huang B, Zheng S, Ma B, Yang Y, Zhang S, Jin L. Using deep learning to predict the outcome of live birth from more than 10,000 embryo data. BMC Pregnancy Childbirth. 2022;22(1):36.

7. Miyagi Y, Habara T, Hirata R, Hayashi N. Feasibility of artificial intelligence for predicting live birth without aneuploidy from a blastocyst image. Reprod Med Biol. 2019;18(2):204-11.

8. Malmsten J, Zaninovic N, Zhan Q, Rosenwaks Z, Shan J. Automated cell stage predictions in early mouse and human embryos using convolutional neural networks. In2019 IEEE EMBS international conference on biomedical & health informatics (BHI) 2019 May 19 (pp. 1-4). IEEE.

9. Liu Z, Huang B, Cui Y, Xu Y, Zhang B, Zhu L, et al. Multi-task deep learning with dynamic programming for embryo early development stage classification from time-lapse videos. IEEE Access. 2019;7: 122153-63.

10. Khan A, Gould S, Salzmann M. Deep convolutional neural networks for human embryonic cell counting. InEuropean conference on computer vision 2016 Oct 8 (pp. 339-348). Springer, Cham.

11. Gingold JA, Ng NH, McAuley J, Lipton Z, Desai N. Predicting embryo morphokinetic annotations from time-lapse videos using convolutional neural networks. Fertil Steril. 2018;110(4):e220.

12. Thirumalaraju P, Kanakasabapathy MK, Bormann CL, Gupta R, Pooniwala R, Kandula H, et al. Evaluation of deep convolutional neural networks in classifying human embryo images based on their morphological quality. Heliyon. 2021;7(2):e06298.

13. VerMilyea M, Hall JM, Diakiw SM, Johnston A, Nguyen T, Perugini D, et al. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. Hum Reprod. 2020;35(4):770-84.

14. Chen TJ, Zheng WL, Liu CH, Huang I, Lai HH, Liu M. Using deep learning with large dataset of microscope images to develop an automated embryo grading system. Fertil Reprod. 2019;1(1):51-6.

15. Louis CM, Erwin A, Galinium M. Automated annotation of in vtrio fertilization time-lapse using deep neural network [dissertation]: Swiss German University; 2020. 105 p.

16. Palermo GD, O'Neill CL, Chow S, Cheung S, Parrella A, Pereira N, et al. Intracytoplasmic sperm injection: state of the art in humans. Reproduction. 2017;154(6):F93-F110.

17. Pribenszky C, Nilselid AM, Montag M. Time-lapse culture with morphokinetic embryo selection improves pregnancy and live birth chances and reduces early pregnancy loss: a meta-analysis. Reprod Biomed Online. 2017;35(5):511-20.

18. Raudonis V, Paulauskaite-Taraseviciene A, Sutiene K, Jonaitis D. Towards the automation of early-stage human embryo development detection. Biomed Eng Online. 2019;18(1):120.

19. O'Shea K, Nash R. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458. 2015 Nov 26.

20. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. Pattern Recognition. 2018;77:354-77.

21. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In2017 international conference on engineering and technology (ICET) 2017 Aug 21 (pp. 1-6). Ieee.

22. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. ArXiv preprint arXiv:1603.04467. 2016 Mar 14.

23. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. {TensorFlow}: A system for {Large-Scale} machine learning. In12th USENIX symposium on operating systems design and implementation (OSDI 16) 2016 (pp. 265-283).

24. Dillon JV, Langmore I, Tran D, Brevdo E, Vasudevan S, Moore D, Patton B, Alemi A, Hoffman M, Saurous RA. Tensorflow distributions. arXiv preprint arXiv:1711.10604. 2017 Nov 28.

25. Chollet F, others. Keras [Internet]. San Francisco: GitHub; 2015 [cited 2021 Apr 16]. Available from: https://github.com/fchollet/keras